Muti-Machine Disruption Challenge by ITU 2023:Ensemble of LSTM and gMLP for Fusion Energy Prediction Final report

Abstract

The objective of this challenge is to develop a disruption prediction model that can be applied universally, using J-TEXT and HL-2A as the current devices and C-Mod as the future device. In order to solve this problem, we trained two models based on LSTM and gMLP respectively to achieve prediction of rupture, and finally we integrated the two models using the GBDT method. This method finally achieved an f1score of 0.782 on C-MOD's rupture prediction.

Introduction

Utilizing nuclear fusion energy via magnetic-confinement tokamaks is one of a few encouraging paths toward future sustainable energy. Along the way, scientists need to learn to avoid plasma disruptions: these sudden and unexpected plasma terminations still represent one of the key challenges for tokamak devices, as their deleterious consequences can damage the whole fusion device and prevent the realization of a burning plasma reactor. Forecasting plasma instabilities and disruptions using first-principle models has demonstrated to be extremely difficult due to the complexity of the problem and the high non-linearity of the system. To date, disruption prediction has been studied through two main approaches: data-driven versus physics-driven (or model-based). On one hand, recent statistical and machine learning approaches based on experimental data have shown attractive results for disruption prediction, even in real-time environments. Different tokamak devices have different operational spaces, spatiotemporal scales for physics events and plasma diagnostics. Therefore, most of these data-driven approaches were developed and optimized specifically for one device and did not show promising cross-device prediction ability. The objective of this challenge is to develop a disruption prediction model that can be applied universally, using J-TEXT and HL-2A as the current devices and C-Mod as the future device.

Data preprocessing

The original data given in this challenge cannot be directly trained. We need to perform data preprocessing before using it to train the model. Regarding data preprocessing, we did the following work: First, according to the official signal correspondence table on the three machines, we extracted the signals in HL-2A and J-TEXT according to the signal names of C-MOD in the table., name all the data in HL-2A and J-TEXT according to the corresponding C-MOD signal names in the table so that other researchers can use these data to train the model.

Then, we align all signals according to their start time by comparing the "start_time" attribute in the properties of each signal. Then, we downsample all signals with a sampling rate of 5000. The purpose of this is to align all the data so that they have the same length for subsequent processing. Finally, we uniformly handle all outliers to ensure that no problems occur during subsequent training.For label division, we use different division methods for the HL-2A and J-TEXT data sets respectively. In HL-2A, we represent the first 100 time steps of the rupture point as 1, and other time steps as 0. In the J-TEXT data set we represent the first 40 time steps of the rupture point as 1 and other time steps as 0.

Proposed Model

Because in the data set processed so far, only the lengths of the signals of each shot are the same, and the lengths of the signals of each shot are not the same. Therefore, before model training, the data was also processed with variable step size sliding window: in most time steps before the front rupture point of the gun that did not rupture or ruptured, a relatively large sliding window step size was used. A small step size is used to divide the sliding window in a relatively short time step before the point. The model can achieve good results by using the classic two-layer LSTM model for training.Among them, we save the model of each epoch, evaluate the model based on the verification loss of the model and analyze the model's predicted waveform for each shot, and finally upload it to the website to judge the true quality of the model.

The data preprocessing method using the gMLP model is basically consistent with the idea of the LSTM model. What is different from the LSTM model is that in the selection of step sizes for most time steps before the rupture point of the non-ruptured cannon and the ruptured cannon, the gMLP model Instead of using a fixed step size, a random step size is used. Moreover, the gMLP model chooses the window size to be 80, which is also different from the LSTM model.We use the tsai library to simplify our model training process.

After we got two models with good results, we used the GBDT method to integrate our models, and then got our final result.Gradient Boosting Decision Trees is an ensemble learning method designed to build a powerful predictive model. The method involves iteratively training decision trees, with each tree focusing on correcting the prediction errors of the preceding one. In each iteration, the model minimizes the residuals through gradient descent, gradually enhancing overall performance. Ultimately, the combination of multiple weak learners forms a robust ensemble model applicable to both regression and classification problems. GBDT excels in handling complex data and nonlinear relationships, making it widely employed in practical applications.